The Genomic Basis of Tumor Regression in Tasmanian Devils (*Sarcophilus harrisii*)

Mark J. Margres[1], Manuel Ruiz-Aravena[2], Rodrigo Hamede[2,3], Menna E. Jones[2], Matthew F. Lawrance[1], Sarah A. Hendricks[4], Austin Patton[1], Brian W. Davis[5,6], Elaine A. Ostrander[6], Hamish McCallum[7], Paul A. Hohenlohe[4], and Andrew Storfer[1,*]

[1]School of Biological Sciences, Washington State University

[2]School of Natural Sciences, University of Tasmania, Hobart, Tasmania, Australia

[3]Centre for Integrative Ecology, Deakin University, Waurn Ponds, Victoria, Australia

[4]Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow

[5]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station

[6]Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland

[7]School of Environment, Griffith University, Nathan, Queensland, Australia

*Corresponding author: E-mail: astorfer@wsu.edu.

Accepted: October 10, 2018

Data deposition: This project has been deposited at the Sequence Read Archive (SRA) under BioProject PRJNA450403 and accessions SRR7015138, SRR7015139, SRR7015141–SRR7015147, and SRR7015149.

Abstract

Understanding the genetic basis of disease-related phenotypes, such as cancer susceptibility, is crucial for the advancement of personalized medicine. Although most cancers are somatic in origin, a small number of transmissible cancers have been documented. Two such cancers have emerged in the Tasmanian devil (*Sarcophilus harrisii*) and now threaten the species with extinction. Recently, cases of natural tumor regression in Tasmanian devils infected with the clonally contagious cancer have been detected. We used whole-genome sequencing and F_{ST}-based approaches to identify the genetic basis of tumor regression by comparing the genomes of seven individuals that underwent tumor regression with those of three infected individuals that did not. We found three highly differentiated candidate genomic regions containing several genes related to immune response and/or cancer risk, indicating that the genomic basis of tumor regression was polygenic. Within these genomic regions, we identified putative regulatory variation in candidate genes but no nonsynonymous variation, suggesting that natural tumor regression may be driven, at least in part, by differential host expression of key loci. Comparative oncology can provide insight into the genetic basis of cancer risk, tumor development, and the pathogenicity of cancer, particularly due to our limited ability to monitor natural, untreated tumor progression in human patients. Our results support the hypothesis that host immune response is necessary for triggering tumor regression, providing candidate genes that may translate to novel treatments in human and nonhuman cancers.

Key words: genotype–phenotype, cancer, genomics, adaptation.

Introduction

Identifying the genetic basis of complex phenotypes is a central goal of modern biology because 1) many fundamental aspects of evolution, such as pleiotropy (Wagner and Zhang 2011), result from the genotype–phenotype relationship (Stadler et al. 2001; Hansen 2006); 2) our ability to ultimately predict evolution depends on understanding these evolutionary dynamics; and 3) characterizing the genetic basis of complex traits, particularly disease-related traits such as cancer risk (Altshuler et al. 2008; Saadatpour et al. 2015), is of critical importance for conservation and medicine (Harold et al. 2009; Daub et al. 2013; Albert and Kruglyak 2015; Campbell et al. 2017; Margres et al. 2018). Indeed, understanding the genotype–phenotype relationship is crucial for

both assessing disease risk and developing personalized prevention and treatment (Katsios and Roukos 2010; Weller et al. 2010; Chin et al. 2011; Freedland 2018). Genomic advances have enabled the identification of loci associated with oncogenesis, tumor progression, and even cancer regression and spontaneous remission (Phan et al. 2003; Yard et al. 2016; Garcia-Hernandez et al. 2017; Tran et al. 2017; Kumar-Sinha and Chinnaiyan 2018). Comparative oncology should provide insight into the genetic basis of cancer risk, tumor development, and the pathogenicity of human cancers, particularly due to our limited ability to monitor natural, untreated tumor progression in human patients for obvious ethical considerations (Khanna et al. 2006; Schiffman and Breen 2015; Millanta et al. 2016).

Most cancers are somatic in origin, but a few transmissible cancers have been documented (reviewed in Ostrander et al. 2016). Such cancers are rare, with the only natural cases found in dogs (canine transmissible venereal tumor or CTVT; Murgia et al. 2006), bivalves (Metzger et al. 2016), and the Tasmanian devil (devil facial tumor disease or DFTD; Pearse and Swift 2006; Pye et al. 2016). Yet within these species, six transmissible cancers have recently been discovered (i.e., several in bivalves and a second in Tasmanian devils; Metzger et al. 2016; Pye et al. 2016), suggesting that these types of cancers may be more common than previously thought. Unlike dogs and bivalves, the Tasmanian devil (*Sarcophilus harrisii*) is threatened with extinction by the emergence of two clonally transmissible cancers (Hawkins et al. 2006; Murchison et al. 2012; Pye et al. 2016; Stammnitz et al. 2018; Storfer et al. 2018). Devil facial tumor disease 1 and 2 (DFT1 and DFT2, respectively) are contagious tumor cell lines and spread via biting, which is common during social interactions (Hamede et al. 2013). DFT1 was first discovered in northeastern Tasmania in 1996 and has since spread more than 80% of the way across the island, causing significant population declines (McCallum 2008; McCallum et al. 2009). DFT2 was discovered in 2014 and is currently restricted to a small region of southeastern Tasmania (Pye et al. 2016). Although comparatively little is known about DFT2 relative to DFT1, the tumors are histologically, cytogenetically, and genetically distinct despite gross morphological similarities and are of independent origin (Pye et al. 2016; Stammnitz et al. 2018).

Low devil genetic diversity (Miller et al. 2011; Hendricks et al. 2017) and silencing of cell surface MHC molecules by DFTD have putatively led to universal susceptibility and ~100% mortality (Siddle et al. 2013). Yet all devil populations continue to persist, suggesting that devils are responding to DFTD (Jones et al. 2008; Lazenby et al. 2018). Recent work in devils has found evidence of an immune response to DFT1 (Pye et al. 2016), rapid evolution in genes related to immune function and cancer risk following disease arrival (Epstein et al. 2016), and few loci of large-effect underlying disease-related phenotypes such as survival following infection (Margres et al. 2018). Collectively, these studies suggest the evolution of resistance and/or tolerance to DFT1 in devils.

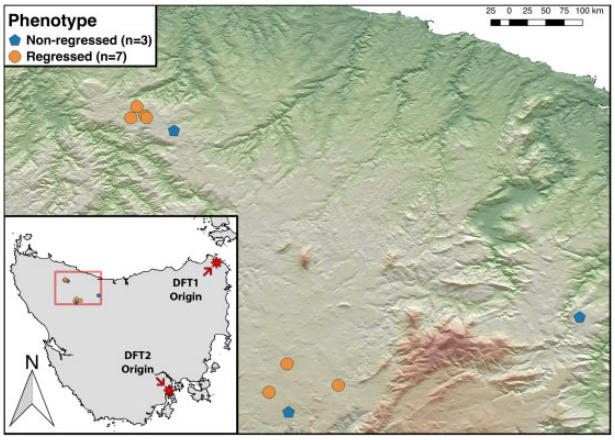

Fig. 1.—Sampling of *Sarcophilus harrisii*. We collected tissue samples from seven individuals that underwent tumor regression and three individuals that did not for whole-genome sequencing. Sampling region as well as the sites of discovery (i.e., putative sites of origin) for DFT1 and DFT2 are indicated in the inset.

Recently, natural (i.e., unmedicated) tumor regression has been detected in northwestern Tasmania (fig. 1; Pye et al. 2016; Wright et al. 2017). In these cases, devils with previously confirmed DFT1 were recaptured with either substantial tumor shrinkage or entire disappearance of the tumor. Recent work on CTVT regression (which, unlike in DFT1, is common) has shown that host response plays a critical role in triggering tumor regression (Frampton et al. 2018), suggesting that DFT1 regression may also be driven by the devil immune system. Wright et al. (2017) recently used whole-genome sequencing (WGS) and a genome-wide association (GWA) approach to compare the genomes of six devils with regressed tumors to the genomes of five devils with nonregressed tumors. Despite significant statistical limitations due to unavoidable sample size limitations (i.e., regression is a rare, novel phenotype), Wright et al. (2017) found evidence that loci involved with angiogenesis may be associated with tumor regression. Although GWA methods are a classical approach for identifying the genomic basis of variation in disease phenotypes, these studies often require several thousand samples to attain appropriate statistical power (Yang et al. 2015). Even for species such as *S. harrisii* with small populations and strong linkage disequilibrium (Epstein et al. 2016), many more than a dozen samples are needed to reliably detect variants of an appreciable effect in a case–control GWA framework (Kardos et al. 2016; Margres et al. 2018). Furthermore, Frampton et al. (2018) showed that tumor regression in CTVT is polygenic and driven by several genes, most notably cyto/chemokines involved with inflammation and interferon signaling. Wright et al.

(2017), however, looked at significant associations for individual SNPs and indels. Although methods for incorporating gene set analyses and other polygenic approaches into a GWA framework exist (Mooney et al. 2014), these models also require many more than a dozen samples (Chatterjee et al. 2013), thus precluding their use for investigating the genetic basis of tumor regression in Tasmanian devils.

We therefore used WGS and F_{ST}-based approaches to compare the genomes of seven individuals that underwent tumor regression (i.e., cases) with those of three infected individuals that did not (i.e., controls) to identify the genetic/polygenic basis of natural tumor regression in Tasmanian devils. We approximately doubled the genome coverage of Wright et al. (2017; 10–15× to 20–30×) and genotyped SNPs, indels, and larger structural variants (SVs) for each individual; the latter were not previously investigated. Genomic instability has repeatedly been shown to be important in many human cancers (e.g., Hartwell 1992; Akagi et al. 2014) as well as in DFT1 and DFT2 (Deakin et al. 2012; Murchison et al. 2012; Pye et al. 2016; Stammnitz et al. 2018). SVs in the human genome, however, can also be associated with disease susceptibility (Tuzun et al. 2005; Manolio et al. 2009) and may have a higher probability, relative to SNPs, of affecting phenotypes (Cooper et al. 2007). We used these three classes of genetic variation to characterize the polygenic basis of tumor regression by identifying large candidate genomic regions and particular biological pathways that differentiated cases and controls. Previous work suggested that loci involved with angiogenesis (Wright et al. 2017), chemokine signaling (Frampton et al. 2018), and cell adhesion (Epstein et al. 2016; Margres et al. 2018) may be important to tumor regression/devil response to DFT1. Frampton et al. (2018) also showed that the differential expression of key genes leads to CTVT regression, and the predominance of expression differentiation underlying rapid, adaptive responses has been documented in humans (Fraser 2013) and other systems (Konczal et al. 2015; Margres et al. 2017). We therefore predicted that loci involved with one or more of these pathways would be highly differentiated between cases and controls, and that this differentiation would be biased toward putative regulatory rather than protein-coding regions. Although transcriptomics is most often used to identify differential gene expression and regulatory variation (Morandin et al. 2016; Breschi et al. 2017; Rokyta et al. 2017), serial sampling of devil tissue pre and postregression was not feasible due to the stochasticity of recapture events. We therefore identified significantly conserved noncoding sequences across devils and other mammals as putative regulatory elements. We then mapped highly differentiated variants to these regions to identify putative regulatory variation underlying the regression phenotype.

Materials and Methods

Sampling and WGS

We sequenced the genomes of ten *S. harrisii* from northwestern Tasmania (fig. 1). All devils were confirmed to have DFT1. Seven of these individuals (six females, one male) exhibited tumor regression (i.e., cases), and three individuals (one female, two males) did not (i.e., controls). Individuals were classified as possessing the regression phenotype if tumor volume decreased ≥15% of the initial tumor volume upon recapture. Genomic DNA was extracted from ear biopsies. Whole-genomes for each individual were sequenced 150 bp paired-end on an Illumina HiSeq X platform to ∼30× coverage. Sequencing was performed at the Northwest Genomics Center at the University of Washington (Seattle, Washington, USA) and GENEWIZ (South Plainfield, New Jersey, USA). All raw data were deposited in the Sequence Read Archive (SRA) under BioProject PRJNA450403 and accessions SRR7015138, SRR7015139, SRR7015141–SRR7015147, and SRR7015149.

Alignments and Variant Calling

We merged raw reads with FLASH2 (Magnoč and Salzberg 2011) and trimmed with Sickle (Joshi and Fass 2011). We then aligned merged and unmerged reads to the reference genome (downloaded from Ensembl January 2016; Murchison et al. 2012) using the BWA-MEM algorithm (Li and Durbin 2009). We identified SNPs and indels using HaplotypeCaller in GATK (McKenna et al. 2010; DePristo et al. 2011) as previously described (Margres et al. 2017). We identified 2,822,764 and 1,194,776 raw SNPs and indels, respectively. Raw variants were filtered in VCFtools (Danecek et al. 2011) and were required to possess a minimum depth value of 15 as well as a minimum genotype quality of 60. Following filtering, the final SNP and indel data sets contained 1, 271, 835 and 281, 046 variants, respectively.

Identifying the Genomic Basis of Tumor Regression

We used the Genotype Phenotype Association Toolkit (GPAT++) from the vcflib package (Garrison 2012) as previously described (Domyan et al. 2016) to identify case–control genomic differentiation. To avoid the high false positive rate inherent to any F_{ST}-based outlier detection approach using small sample sizes (Lotterhos and Whitlock 2015; Verity et al. 2017), we used the segmentFst function from GPAT++ to identify candidate genomic regions with continuous, elevated F_{ST} values relative to the surrounding genome (Domyan et al. 2016); we were much less likely to find highly differentiated, large genomic regions encompassing many variants through sampling error than single variants spread across the genome. We used this approach to identify candidate genomic regions using the SNP and indel data independently. Differentiation was visualized as a Manhattan plot using the pFst function from GPAT++ (Garrison 2012). pFst

uses a likelihood ratio test for allele frequency differences between case–control groups while incorporating genotype likelihood information to account for potential sequencing error. SNPs and indels in candidate genomic regions were characterized using variant effect predictor (VEP; McLaren et al. 2016) and the reference genome (downloaded from Ensembl January 2016; Murchison et al. 2012). Each variant was classified as intergenic, intronic, synonymous, nonsynonymous, being in an untranslated region (UTR), or upstream or downstream of a particular gene (i.e., ≤5 kb of the coding sequence). We used GeneCards (www.genecards.org; Stelzer et al. 2016) to identify putative gene functions.

Identifying Structural Variation

We used Manta (Chen et al. 2016) to identify SVs in the ten individual genomes. Manta uses paired and split-read evidence to identify large indels, inversions, duplications, and translocations. We ran a joint diploid sample analysis across all individuals and identified 79,453 variants that passed default Manta filters. We then calculated Weir and Cockerham's F_{ST} (θ; Weir and Cockerham 1984) to compare differentiation between cases and controls for each variant. Because the reference genome has only been assembled into ~36,000 scaffolds (Murchison et al. 2012) and contains many missing base pairs, more than half of the identified large indels mapped to these regions and were, therefore, false positives. We manually removed variants that mapped to regions of the reference genome with ≥50% N's. We then took the remaining top 0.1% of SVs (segmentFst did not identify any differentiated genomic regions in this data set; data not shown) and required a minimum phred-scaled quality score of 999. This conservative approach identified 35 top SVs that passed all filters.

Detecting Polygenic Selection

Because conventional methods such as integrated haplotype score and derived intra-allelic nucleotide diversity have little power to detect signatures of polygenic adaptation (i.e., small allele frequency changes at many loci; Fagny et al. 2014), we used Weir and Cockerham's F_{ST} (θ; Weir and Cockerham 1984) to compare differentiation between cases and controls for six pathways (Power et al. 2017) relevant to cancer risk and DFT1 response. We used one-sided Mann–Whitney *U* tests to compare the level of case–control differentiation in a given biological pathway to that of the rest of the genome as previously described (Hsieh et al. 2016). Pathways exhibiting greater differentiation than the genomic background may be evolving under positive selection. We examined six biological pathways previously shown to be associated with DFTD response and/or cancer-related traits: Cell adhesion (Epstein et al. 2016; Margres et al. 2018; $n = 2442$ genes), apoptosis ($n = 115$; Epstein et al. 2016), cancer ($n = 229$; Rosenberg et al. 2008), graft-versus-host ($n = 13$; Rosenberg et al. 2008), chemokine signaling ($n = 79$; Frampton et al. 2018), and vascular endothelial growth factor signaling ($n = 40$; Wright et al. 2017). Genes for each pathway were identified in the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto 2000). All variants within 5 kb of the coding-region of a gene were listed as a pathway variant. SNPs and indels were analyzed independently. All genes and pathways are provided in supplementary table S1, Supplementary Material online.

Identifying Regulatory Regions from Whole-Genome Sequences

To determine if candidate variants occurred in putative regulatory regions, we used gVISTA (Couronne et al. 2003) and RankVISTA to identify conserved noncoding sequences (CNS) across devils and humans. RankVISTA uses the Gumby algorithm to estimate neutral evolutionary rates from nonexonic regions in the alignment and identify statistically significant segments that evolve more slowly than the background; phylogenetically weighted log-odds conservation scores of conserved segments are translated to *P*-values using Karlin-Altschul statistics (Martin 2004). The human genome (NCBI v36.1, hg18, March 2006) was used as the reference for all analyses. Identified CNS represent putative regulatory elements. We used a RankVISTA threshold of 0.5, a calculation window of 100 bp, and limited these comparisons to the candidate genomic regions identified in our segmentFst analyses described above. Candidate genomic regions that occurred on scaffolds that did not contain annotated coding sequences were excluded. For the shared SNP/indel candidate regions, we took the earliest starting and latest stopping positions to analyze the largest possible region. Because gVISTA cannot analyze sequences >300 kb, we cut the larger genomic regions into 300 kb blocks for these analyses. Significantly conserved intronic and intergenic sequences were identified as putative regulatory elements. Results were visualized in VISTA (Dubchak et al. 2000; Mayor et al. 2000; Frazer et al. 2004) using six nondevil mammal genomes: Humans, mouse, rat, dog, horse, and rhesus monkey. SNPs and indels were then mapped to the identified CNS and annotated as putative regulatory variants.

Results

Candidate Genomic Regions Underlying Tumor Regression in Tasmanian Devils

We identified 11 highly differentiated genomic regions between cases and controls using the SNP data (fig. 2*A*). These candidate regions occurred on ten scaffolds across 5 chromosomes and contained 22 annotated genes, 10 of which had a putative immune and/or cancer-related function

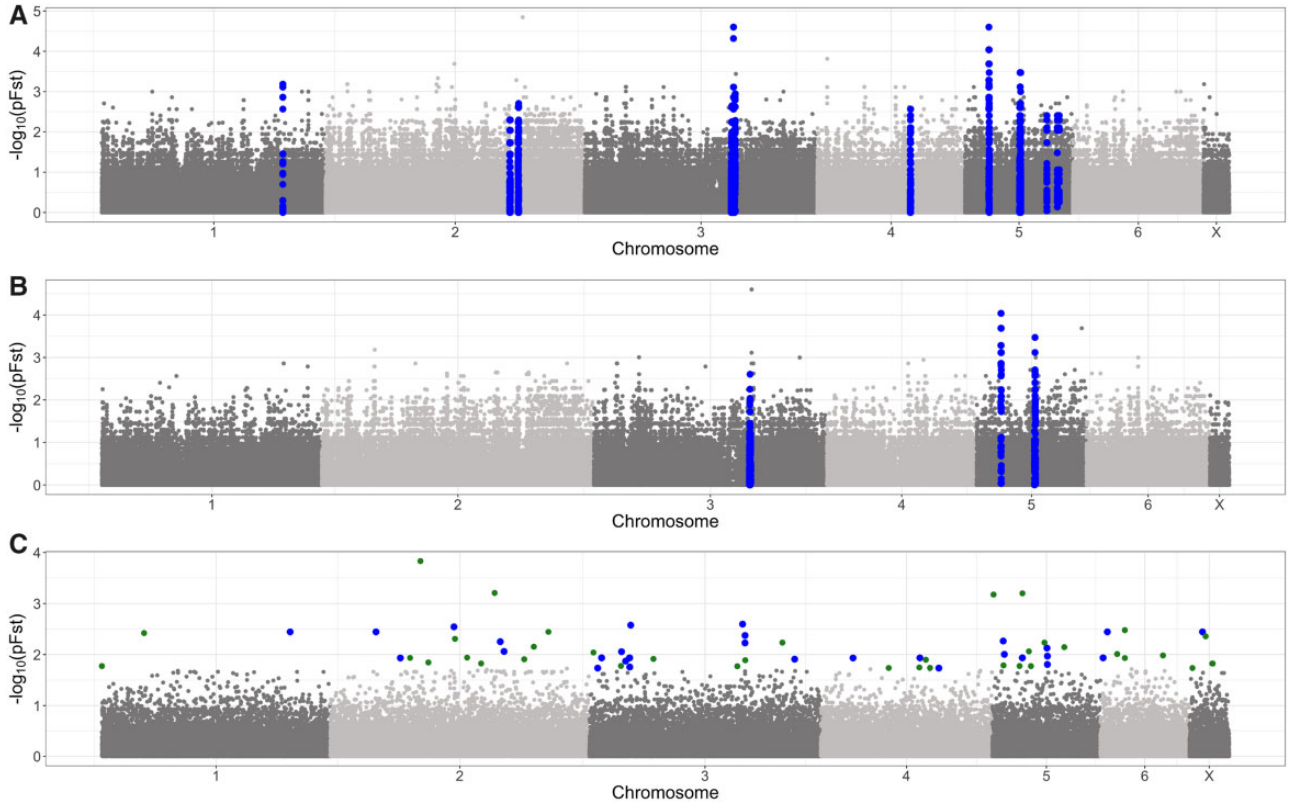

Fig. 2.—Manhattan plots showing differentiation between cases and controls for (*A*) SNPs, (*B*) indels, and (*C*) structural variants. Blue points in (*A*) and (*B*) indicate scaffolds that were identified as highly differentiated candidate genomic regions by segmentFst. The three candidate indel scaffolds in (*B*) are the same scaffolds as in (*A*). Because the *x* axis is variant index rather than true genomic position (i.e., the 36,000 scaffolds of the reference genome makes robust physical mapping problematic), the scaffolds only appear slightly shifted. (*C*) Blue points indicate empirical outliers that passed all filters (see Materials and Methods). Green points indicate empirical outliers that did not pass all filters. The *y* axis in all panels represents the negative log of the pFst statistic.

(table 1). Chromosome 6 and the X chromosome did not possess any candidate regions.

We identified three highly differentiated genomic regions between cases and controls using the indel data (fig. 2*B*). These candidate regions occurred on three scaffolds across chromosomes 3 and 5. All three candidate regions occurred on scaffolds also identified in the SNP-based analysis. The SNP candidate genomic region on chromosome 3 was completely encompassed by the larger indel candidate region. The candidate indel region on chromosome 5 scaffold GL861622 was completely encompassed by the larger SNP candidate region, and the indel region on chromosome 5 scaffold GL861682 overlapped both candidate SNP regions identified on this scaffold (fig. 2*A* and *B*; tables 1 and 2). The 3 candidate indel regions contained 12 annotated genes, 7 of which had a putative immune and/or cancer-related function (table 2). All seven genes in the candidate indel regions on chromosome 5 were identified in the SNP-based analysis, but four of the five genes identified in the candidate indel region on chromosome 3 were not identified in the SNP-based analysis (only *MED12L* was shared across the SNP and indel candidate regions on chromosome 3; tables 1 and 2). Three of the four genes identified only in the candidate indel region on chromosome 3 had a putative immune and/or cancer-related function (table 2).

Candidate SNPs Underlying Tumor Regression in Tasmanian Devils

We identified 1148 SNPs in the 11 SNP-based candidate genomic regions described above (supplementary table S2, Supplementary Material online), 585 of which we considered highly differentiated ($F_{ST} \geq 0.500$). Most of the highly differentiated SNPs were intergenic (69.7%) and intronic (26.8%; fig. 3*A* and table 1). Nineteen (~3.3%) were up or downstream of a coding-region, a single highly differentiated SNP was synonymous, and no nonsynonymous candidates were identified (fig. 3*A* and table 1).

Forty-nine candidate SNPs occurred in or near loci with putative cancer-related functions (table 1). We identified two highly differentiated SNPs in the introns of *WDR48* ($0.560 \leq F_{ST} \leq 0.577$), a putative tumor suppressor, and

Table 1
Candidate Genomic Regions Identified in the SNP-Based segmentFst Analysis with Corresponding Genes and Candidate SNPs

Chromsome	Scaffold	Start–End	Gene	Putative Function	Candidate SNPs
1	GL834900	59090–264617	Protein phosphatase 2, regulatory subunit B, delta (*PPP2R2D*)[a]	Cell growth/division control; Activated TLR4 signaling[a]	–
			BCL2/adenovirus E1B 19 kDa interacting protein 3 (*BNIP3*)[a]	Apoptosis and authophagy[a]	–
			Janus kinase and microtubule interacting protein 3 (*JAKMIP3*)	Microtubule and kinase binding	13 intronic
2	GL841602	886632–992998	Xylulokinase (*XYLB*)	Metabolism	8 intronic 1 downstream
2	GL841626	476121–923043	Sodium channel, voltage-gated, type XI, alpha subunit (*SCN11A*)	Action potential and pain signaling	4 upstream
			WD repeat domain 48 (*WDR48*)[a]	Deubiquitinating complex regulator; Potential tumor suppressor[a]	2 intronic
			Golgi reassembly stacking protein 1 (*GORASP1*)	Protein sorting and modification	1 upstream
			Tetratricopeptide repeat domain 21A (*TTC21A*)	Protein localization	–
			Xin actin-binding repeat containing 1 (*XIRP1*)[a]	Cellular junction function; leukemia candidate gene[a]	–
			Solute carrier family 22, member 14 (*SLC22A14*)	Molecule transport	–
			Solute carrier family 22, member 13 (*SLC22A13*)	Molecular transport and metabolism	–
			Oxidative stress responsive 1 (*OXSR1*)[a]	Kinase response to environmental stress[a]	–
3	GL849905	3117969–3162680	Mediator complex subunit 12-like (*MED12L*)	Transcriptional coactivation	–
4	GL856968	1213224–1266653	–	–	–
5	GL861622	8458–1345777	Lymphoid-restricted membrane protein (*LRMP*)[a]	Peptide delivery to MHC class I; related to b-cell lymphomas[a]	–
			Cancer susceptibility candidate 1 (*CASC1*)[a]	Microtubule and tubulin binding; related to respiratory neoplasms[a]	–
			LYR motif containing 5 (*LYRM5*)	Protein synthesis	–
			Kirsten rat sarcoma viral oncogene homolog (*KRAS*)[a]	Member of small GTPase superfamily; implicated in multiple carcinomas[a]	1 intronic
			Lamin tail domain containing 1 (*LMNTD1*)[a]	Structural activity; related to respiratory neoplasms[a]	30 intronic 4 upstream 4 downstream
			Ras association domain family member 8 (*RASSF8*)[a]	Tumor suppressor protein; epithelial cell migration[a]	–
			Sarcospan (*SSPN*)	Structural support	–
5	GL861682	1183188–1317129	Glutamate decarboxylase 2 (*GAD2*)	L-glutamic acid catalysis	87 intronic 2 upstream 1 synonymous
5	GL861682	1500464–1544340	Myosin IIIA (*MYO3A*)[a]	Hearing; associated with bladder/colon cancer[a]	7 intronic 1 upstream
5	GL861847	655–43394	–	–	–
5	GL862178	7209–28728	ENSSHAG00000000783	Unannotated coding sequence	9 intronic 2 downstream
5	GL862286	963–11539	–	–	–

[a]Putative immune or cancer-related function. Candidate SNPs were variants with $F_{ST} \geq 0.5$.

Table 2
Candidate Genomic Regions Identified in the Indel-Based segmentFst Analysis with Corresponding Genes and Candidate Indels

Chromsome	Scaffold	Start–End	Gene	Putative Function	Candidate Indels
3	GL849905	2931754–3563755	*MED12L*	Transcriptional coactivation	6 intronic
					1 upstream
					4 intronic[b]
			G protein-coupled receptor 87 (*GPR87*)[a]	Regulated by *p53*; overexpressed in carcinomas[a]	1 intronic[b]
					3 downstream[b]
			Purinergic receptor P2Y, G-protein coupled, 14 (*P2RY14*)[a]	Receptor activity; immune and neuroimmune function[a]	–
			G protein-coupled receptor 171 (*GPR171*)	Ligand-binding receptor	–
			Siah E3 ubiquitin protein ligase 2 (*SIAH2*)[a]	Hypoxia response; involved in apoptosis and tumor suppression[a]	–
5	GL861622	159894–1238762	*CASC1*[a]	Microtubule and tubulin binding; related to respiratory neoplasms[a]	1 intronic
			LYRM5	Protein synthesis	–
			KRAS[a]	Small GTPase superfamily member; implicated in multiple carcinomas[a]	–
			LMNTD1[a]	Structural activity; related to respiratory neoplasms[a]	7 intronic
					1 upstream
					1 downstream
			RASSF8[a]	Tumor suppressor protein; epithelial cell migration[a]	1 intronic
5	GL861682	1186331–1550521	*GAD2*	L-glutamic acid catalysis	11 intronic
			MYO3A[a]	Hearing; associated with bladder/colon cancer[a]	21 intronic
					1 upstream

[a]Putative immune or cancer-related function.
[b]Indels found in overlapping open-reading frames and, therefore, with ambiguous designations. Candidate indels were variants with $F_{ST} \geq 0.5$. Abbreviated gene names were defined in table 1.

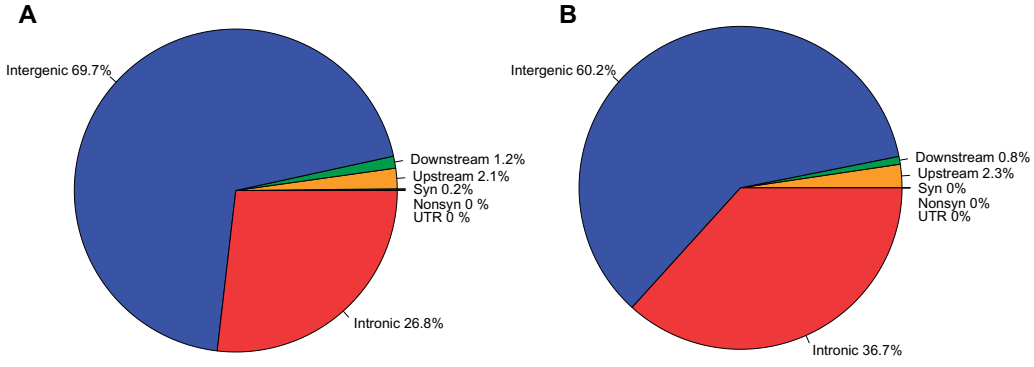

Fig. 3.—Pie charts depicting the relative frequencies of different mutational classes among the most differentiated (*A*) SNPs ($n = 585$) and (*B*) indels ($n = 128$) in the candidate genomic regions from figure 2. Nonsyn, nonsynonymous; Syn, synonymous; UTR, untranslated region.

a top SNP in the intron of *KRAS* (F_{ST}=0.765), a gene implicated in multiple carcinomas. We also identified 38 highly differentiated SNPs ($0.566 \leq F_{ST} \leq 0.790$) in or near *LMNTD1* and eight candidate SNPs ($0.500 \leq F_{ST} \leq 0.536$) in or near *MYO3A*. *LMNTD1* is related to respiratory neoplasms, and although *MYO3A* primarily functions in hearing, modifications in *MYO3A* were found to be associated with bladder and colon cancer (see Discussion; Lascorz et al. 2010; Chung et al. 2011).

Candidate Indels Underlying Tumor Regression in Tasmanian Devils

We identified 197 indels in the 3 indel-based candidate genomic regions (supplementary table S3, Supplementary Material online), 132 of which were highly differentiated ($F_{ST} \geq 0.500$). Excluding the four ambiguous highly differentiated indels identified in the overlapping gene region on chromosome 3 (see below), most of the highly differentiated indels were intergenic (60.2%) and intronic (36.7%; fig. 3*B* and table 2). Four indels (~3.1%) were up or downstream of a coding-region. As expected, no indels were identified in the coding-regions or UTRs of genes (fig. 3*B* and table 2).

For the candidate region on chromosome 3 (fig. 2*B* and table 2), we identified four variants with ambiguous designations because *MED12L* (scaffold position 2786476–3136515) and *GPR87* (scaffold position 2912235–2934071) are overlapping genes. Overlapping genes have been documented in vertebrate genomes, including humans and other mammals (e.g., Makalowska et al. 2005). Based on the annotation of the reference *S. harrisii* genome (downloaded from Ensembl January 2016; Murchison et al. 2012), *GPR87* is nested within *MED12L*. We identified one highly differentiated variant (F_{ST}=0.644) in intronic regions of both genes as well as three variants that occurred in an intronic region of *MED12L* and downstream of *GPR87* ($0.571 \leq F_{ST} \leq 0.644$). *MED12L* functions in transcriptional coactivation, and *GPR87* is regulated by *p53* (table 2).

Several other candidate indels also occurred in or near loci with putative cancer-related functions (table 2). We identified seven intronic candidates ($0.804 \leq F_{ST} \leq 0.864$) as well as individual up- (F_{ST}=0.884) and downstream candidates (F_{ST}=0.831) for *LMNTD1*. We also identified candidate indels in the introns of *RASSF8* (F_{ST}=0.686) and *CASC1* (F_{ST}=0.556). *RASSF8* is a putative tumor suppressor, and *CASC1* has also been associated with neoplasms, similar to *LMNTD1* (table 2). Twenty-one intronic ($0.516 \leq F_{ST} \leq 0.703$) and an upstream candidate indel (F_{ST}=0.703) were found for *MYO3A* (table 2); *MYO3A* also possessed several candidate SNPs (table 1) and has been implicated in cancer susceptibility (Lascorz et al. 2010; Chung et al. 2011).

Candidate SVs Underlying Tumor Regression in Tasmanian Devils

We identified 35 highly differentiated putative SVs that passed all filters (fig. 2*C*). Of these 35 candidate SVs, 21 were intergenic, 11 were found in the introns of 10 genes, 1 variant was upstream of a gene, and 2 variants were downstream of a single gene. Of the 14 variants in or near genes, we identified 6 insertions (58–584 bp), 2 deletions (95–614 bp), and 6 putative translocations (table 3). Both deletions and all but one insertion occurred in introns; one insertion was upstream of a coding-sequence. Of the six putative translocations, four occurred in introns and two occurred in downstream regions, although both putative downstream translocations also involved introns of another gene (i.e., we genotyped six highly differentiated translocations, but two overlapped, leaving four candidate translocations; see below; table 3).

Six of the 12 genes associated with a candidate SV were implicated in immune-response and/or cancer-risk (table 3). We identified insertions in the introns of *CSRNP3*, *MAP3K5*, and *SEC31A*; the former two genes are involved with apoptosis, and the latter is associated with inflammatory myofibroblastic tumors. We also found putative translocations involving three candidate genes (table 3). We identified a putative translocation between the intron of *TRIM33*, which is involved with the ligation of E3 ubiquitin-protein and is related to carcinomas, with an intergenic region on chromosome three. We also identified two putative translocations between the downstream regions of *LMNTD1* and the intronic regions of *TNKS*. *LMNTD1* has been associated with respiratory neoplasms and was identified as a candidate gene in both the SNP (table 1) and indel (table 2) analyses. *TNKS* is involved with Wnt signaling and associated with various forms of human cancer (table 3).

Detecting the Polygenic Basis of the Regression Phenotype

F_{ST} distributions for all six biological pathways for SNPs ($0.19 \leq P \leq 1.00$) and indels ($0.83 \leq P \leq 1.00$) were nearly identical to that of the rest of the genome, indicating an absence of detectable polygenic selection in these particular pathways (supplementary fig. S1, Supplementary Material online).

Identification of Putative Regulatory Elements in the Candidate Genomic Regions

We identified CNS in six of the seven candidate genomics regions analyzed; the first candidate region on chromosome 2 (scaffold GL841602) did not return a match. Although we identified between 1 and 52 CNS per candidate genomic region, only one candidate genomic region, scaffold GL861682 on chromosome 5, possessed variants in the identified CNS (fig. 4).

On chromosome 5 scaffold GL861682, we identified 23 conserved intronic sequences, all of which were introns for *MYO3A*, and a single conserved intergenic sequence; these regions were 102–541 bp and 67.5–85.4% similar at the sequence level. Two of the conserved *MYO3A* intronic sequences, which represent putative regulatory elements, possessed SNPs that were highly differentiated between cases and controls (fig. 4). The first candidate regulatory intronic region (position 1438502–1438639, 73.6% similarity to the human genome) possessed a highly differentiated SNP at position 1438555 (F_{ST}=0.536). The second candidate regulatory intronic region (position 1498234–1498453, 73.3% similarity to the human genome) possessed two candidate SNPs at positions 1498270 (F_{ST}=0.536) and 1498411 (F_{ST}=0.491).

Table 3
Candidate Genes Identified in the Structural Variant Analysis

Chromsome	Scaffold	Position	Type	Gene	Putative Function
2	GL841465	252851	58 bp insertion intron	DTW domain containing 2 (*DTWD2*)	Mitochondrion protein targeting
	GL841584	1176582	95 bp deletion intron	BMP binding endothelial regulator (*BMPER*)	Bone morphogenetic protein inhibition
3	GL849630	133609	584 bp insertion upstream	Solute carrier family 38 member 11 (*SLC38A11*)	Amino acid transport
	GL849631	677692	278 bp insertion intron	Cysteine-serine-rich nuclear protein 3 (*CSRNP3*)[a]	Transcriptional activator regulator; apoptosis[a]
4	GL856953	1140408	247 bp insertion intron	Mitogen-activated protein kinase kinase kinase 5 (*MAP3K5*)[a]	Apoptosis signal transduction[a]
	GL856993	210323	Translocation of intron with intergenic region on chromosome 3 (GL849549)	Tripartite motif containing 33 (*TRIM33*)[a]	E3 ubiquitin-protein ligase; related to carcinomas[a]
5	GL861591	1045356	614 bp deletion intron	ENTH domain containing 1 (*ENTHD1*)	Unknown
	GL861622	511673	Translocation of downstream region with intron of tankyrase (*TNKS*) on chromosome 6 (see below)	*LMNTD1*[a]	Structural activity; related to respiratory neoplasms[a]
		511686	Translocation of downstream region with intron of *TNKS* on chromosome 6 (see below)	*LMNTD1*[a]	Structural activity; related to respiratory neoplasms[a]
	GL861672	1578145	69 bp insertion intron	GTP binding protein 4 (*GTPBP4*)	Cell signaling
6	GL864736	1234425	Translocation of intron with downstream region of *LMNTD1* on chromosome 5 (see above)	*TNKS*[a]	Wnt signaling; target of cancer treatment[a]
		1234867	Translocation of intron with downstream region of *LMNTD1* on chromosome 5 (see above)	*TNKS*[a]	Wnt signaling; target of cancer treatment[a]
	GL864753	78910	395 bp insertion intron	SEC31 homolog A (*SEC31A*)[a]	Vesicle budding; related to inflammatory myofibroblastic tumors[a]
X	GL867607	3118283	Translocation of intron with intergenic region on chromosome 1 (GL835007)	Dachshund family transcription factor 2 (*DACH2*)	Transcription factor

[a]Putative immune or cancer-related function. Abbreviated gene names were defined in tables 1 and 2.

Discussion

We predicted that the genomic basis of tumor regression in Tasmanian devils would be 1) polygenic; 2) involve specific biological pathways related to cancer risk and DFT1 response; and 3) be biased toward putative regulatory rather than protein-coding regions. We found three highly differentiated candidate genomic regions, indicating that the genomic basis of tumor regression in Tasmanian devils was polygenic. We did not detect signatures of polygenic selection in the six pathways investigated, suggesting that the genetic basis of tumor regression is either associated with the candidate genomic regions discussed above and/or other biological pathways. We identified putative regulatory variation in genes related to immune response and/or cancer risk but no nonsynonymous variation, suggesting that natural DFT1 regression may be driven, at least in part, by differential host expression of key candidate genes. We discuss each outcome in detail below.

First, we identified three highly differentiated candidate regions across chromosomes 3 and 5. Although not all variants in these candidate regions were necessarily associated with the regression phenotype and may simply be linked to

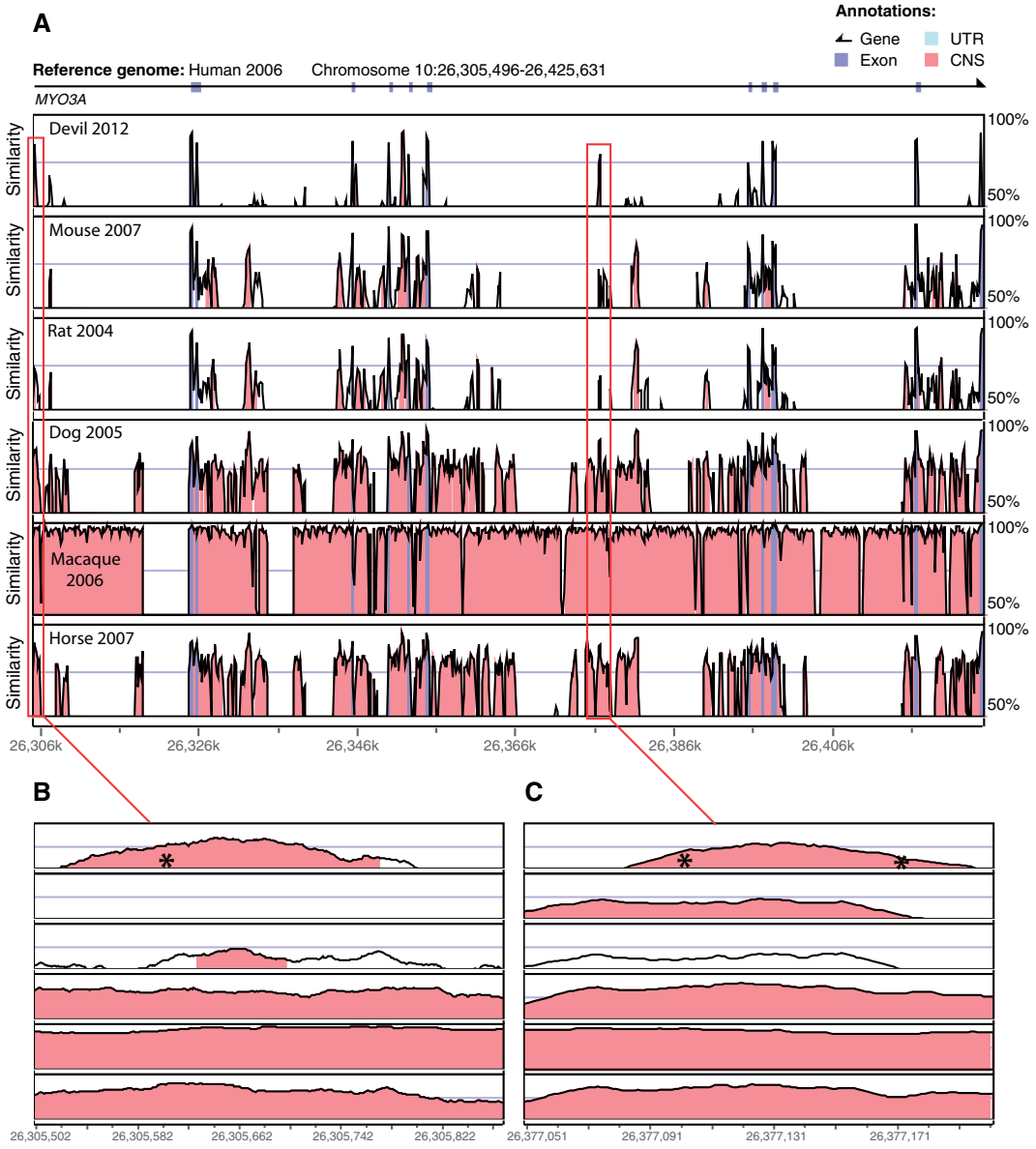

Fig. 4.—VISTA plot showing conserved sequences across devils and six other mammals for a candidate genomic region containing *MYO3A*. (*A*) We plotted 1,400,000–1,500,000 on devil chromosome 5 scaffold GL861682. Two conserved intronic sequences (*B* and *C*), indicated by the red boxes in (*A*), possessed candidate SNPs that were highly differentiated ($F_{ST} \geq 0.491$) between cases and controls, suggesting that these SNPs may lead to regulatory variation in *MYO3A*. Genomic position in the reference human genome is listed on the *x* axis in all panels. Percent similarity is shown on the *y* axis where the blue line indicates 70% identity in all panels. Asterisks represent approximate location of candidate SNPs. CNS, conserved noncoding sequence; UTR, untranslated region.

one or a few large-effect variants, the strong signal of differentiation across multiple genomic regions suggested that the regression phenotype was indeed polygenic. These regions were independently identified in the SNP- and indel-based analyses and contained candidate genes such as *GPR87*, *CASC1*, *KRAS*, *RASSF8*, and *LMNTD1*, all of which contained one or more candidate variants. *LMNTD1* was identified as a candidate gene in the SNP-, indel-, and SV-based analyses, although the SV-based results must be interpreted with caution; the fragmented state of the reference genome (i.e., ~36,000 scaffolds) made interpretation difficult.

Second, we did not detect signatures of polygenic selection in the six pathways investigated. Given that tumor regressions were first detected in 2009 (Pye et al. 2016; Wright et al. 2017) and that first-step mutations are often of large-effect (Rokyta et al. 2005), minor allele frequency changes at many loci may not have had enough time to occur (e.g., ~4–5

generations assuming a 2-year generation time; McCallum et al. 2009; Hamede et al. 2015). The polygenic basis of DFT1 regression may therefore be in the form of a few, large-effect variants rather than many variants of small-effect, consistent with other work investigating the genetic basis of DFT1-related phenotypes in devils (e.g., survival following infection; Margres et al. 2018).

Third, none of the candidate variants were nonsynonymous, and only a single candidate synonymous SNP was identified, suggesting that differential expression of key candidate genes in devils may play a role in tumor regression. Indeed, Wright et al. (2017) did not find any candidate SNPs in coding-sequences using GWA approaches investigating the same trait, and CTVT regression in dogs also appears to be driven, at least in part, by differential gene expression (Frampton et al. 2018). Noncoding variation has also repeatedly been linked to tumorigenesis and other diseases in humans (e.g., Mathelier et al. 2015). Bias toward differential expression rather than nonsynonymous variation, however, is not specific to tumor regression or cancer-related phenotypes. Several studies have suggested that adaptive expression differentiation may be more likely than adaptive changes to coding regions, at least for rapid adaptation in complex phenotypes, because more mutational mechanisms exist for altering expression levels than for protein-coding sequences (Fraser 2013; Konczal et al. 2015; Margres et al. 2017). Our results are consistent with these expectations.

We also found three candidate SNPs in two significantly conserved intronic sequences in *MYO3A*. Introns can function as regulatory elements (e.g., Majewski and Ott 2002; Wei et al. 2006), and the conservation of these intronic regions across ~160 Myr of evolution (marsupial-placental split; Luo et al. 2011) indicated that these sequences may have an important regulatory function. These three candidate SNPs, therefore, may lead to differential expression of *MYO3A* between cases and controls. Although *MYO3A* primarily functions in transferase/kinase activity and plays a key role in hearing in humans, *MYO3A* is highly methylated in bladder tumors (Chung et al. 2011). A GWA study also found that an SNP in the intron of *MYO3A* was associated with an increased risk for colon cancer in humans (Lascorz et al. 2010), similar to our results (although the candidate colon cancer SNP, rs11014993, occurred in a different intron). Future transcriptomic and functional genomic studies of serially sampled devils with and without the regression phenotype are needed to determine if *MYO3A*, along with the other candidate genes, are differentially expressed to determine what role these genes may have in tumor regression.

Despite substantial statistical limitations due to the rarity of the regression phenotype and therefore unavoidable small sample sizes, Wright et al. (2017) used a GWA framework and found evidence that SNPs in the introns of loci involved with angiogenesis, particularly *PAX3*, may be associated with tumor regression. *PAX3* occurs on chromosome 3 but on a different scaffold than the candidate region we identified on the same chromosome. We did identify variants in the introns as well as up and downstream of *PAX3*, but none of these variants were highly differentiated. The top variant for *PAX3* was an intronic SNP (F_{ST}=0.307), but mean F_{ST} for *PAX3* intronic SNPs was ~0.026 and all F_{ST} estimates for indels near *PAX3* were ~0. *PAX3*, however, is involved in angiogenesis through regulation of bone morphogenic proteins (BMP). We found a 95-bp deletion in the intron of *BMPER* on chromosome 2, indicating that angiogenesis may indeed play an important role in natural tumor regression (although our polygenic selection analysis suggested otherwise).

Comparative oncology should lead to advances in both human and animal health (Schiffman and Breen 2015), and tumor regression is not specific to transmissible cancers. Spontaneous tumor regression, albeit rare (~1 in 60–100,000 cases; Missotten et al. 2008), has been documented in human cancers (Sengupta et al. 2010). One such cancer is Merkel cell carcinoma (MCC), a rare type of skin cancer that often appears on the face, head, or neck in the form of a nodule (Goessling et al. 2002). MCC is of neuroendocrine origin (Goessling et al. 2002) and may be associated with Merkel cell polyomavirus, somewhat similar to DFT1's neuroectoderm/Schwann cell origin and transmissible nature (Murchison et al. 2010). The first spontaneous regression of MCC was documented in 1986 (O'Rourke and Bell 1986), and regression has occurred at least 22 times since. The mechanism of MCC regression is unclear but may be driven by a T-cell-mediated immune response (Connelly et al. 2000; Pang et al. 2015).

Understanding the genetic basis of regression in DFT1 and other nonhuman cancers may enable the identification of general mechanisms underlying tumor regression in MCC and other human cancers. Indeed, ethical considerations often preclude investigating natural, untreated tumor progression in human cancers, highlighting the importance of comparative oncological studies investigating the genetic basis of general tumor phenotypes such as regression. The nonlinear relationship between genotype and cancer-related traits has been labeled as one of the biggest challenges in biomedical sciences (Katsios and Roukos 2010), and comparative oncology should advance our knowledge of cancer progression, tumor regression, and cancer-normal cell interactions and coevolution (Schiffman and Breen 2015). We identified specific, putative regulatory variants in or near genes related to immune response and/or cancer risk that differentiated cases from controls, indicating that natural tumor regression may be driven, at least in part, by differential host expression of key genes. Our results support the concept that host immune response is necessary for triggering tumor regression (Frampton et al. 2018), providing candidate genes and mechanisms that may translate to novel treatments in human and nonhuman cancers.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was funded by NIH Grant R01-GM126563-01 and NSF Grant DEB 1316549 to A.S., P.A.H., H.M., and M.E.J. as part of the joint NIH-NSF-USDA Ecology and Evolution of Infectious Diseases program, the Australian Research Council (ARC) Future Fellowship (FT100100250) to M.J., ARC Large Grants (A00000162) to M.J., Linkage (LP0561120) to M.J. and H.M., Discovery (DP110102656) to M.J. and H.M., and DECRA (170101116) to R.H. Animal use was approved by the IACUC at Washington State University (ASAF#04392), the University of Tasmania Animal Ethics Committee (A0008588, A0010296, A0011696, A0013326, A0015835), and the Department of Primary Industries, Parks, Water and Environment Animal Ethics Committee. Additional support was provided by NIH P30 GM103324. We thank Allison Fuiten for advice on the VISTA analysis.

Literature Cited

Akagi K, et al. 2014. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res. 24(2):185–199.
Albert F, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 16(4):197–212.
Altshuler D, Daly M, Lander E. 2008. Genetic mapping in human disease. Science 322(5903):881–888.
Breschi A, Gingeras T, Guigo R. 2017. Comparative transcriptomics in human and mouse. Nat Rev Genet. 18(7):425–440.
Campbell CS, Adams CE, Bean CW, Parsons KJ. 2017. Conservation evo-devo: preserving biodiversity by understanding its origins. TREE 32(10):746–759.
Chatterjee N, et al. 2013. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 45(4):400–405.
Chen X, et al. 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32(8):1220–1222.
Chin L, Andersen J, Futreal P. 2011. Cancer genomics: from discovery science to personalized medicine. Nat Med. 17(3):297–303.
Chung W, et al. 2011. Detection of bladder cancer using novel DNA methylation biomarkers in urine sediments. Cancer Epidemiol Biomarkers Prev 20(7):1483–1491.
Connelly T, Cribier B, Brown T, Yanguas I. 2000. Complete spontaneous regression of Merkel cell carcinoma: a review of the 10 reported cases. Dermatol Surg. 26(9):853–856.
Cooper G, Nickerson D, Eichler E. 2007. Mutational and selective effects on copy-number variants in the human genome. Nat Genet. 39(7 Suppl):S22–S29.
Couronne O, et al. 2003. Strategies and tools for whole-genome alignments. Genome Res. 13(1):73–80.
Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.
Daub J, et al. 2013. Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol. 30(7):1544–1558.
Deakin J, et al. 2012. Genomic restructuring in the Tasmanian devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. PLoS Genet. 8(2):e1002483.
DePristo M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5):491–498.
Domyan E, et al. 2016. Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. eLife 5:e12115.
Dubchak I, et al. 2000. Active conservation of noncoding sequences revealed by 3-way species comparisons. Genome Res. 10(9):1304–1306.
Epstein B, et al. 2016. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat Commun. 7:12684.
Fagny M, et al. 2014. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. Mol Biol Evol. 31(7):1850–1868.
Frampton D, et al. 2018. Molecular signatures of regression of the canine transmissible venereal tumor. Cancer Cell 33(4):620–633.
Fraser H. 2013. Gene expression drives local adaptation in humans. Genome Res. 23(7):1089–1096.
Frazer K, Pachter L, Poliakov A, Rubin E, Dubchak I. 2004. Vista: computational tools for comparative genomics. Nucleic Acids Res. 32(Web Server issue):W273–W279.
Freedland S. 2018. Prostate cancer: race and prostate cancer personalized medicine: the future. Nat Rev Urol. 15(4):207–208.
Garcia-Hernandez M, et al. 2017. A unique cellular and molecular microenvironment is present in tertiary lymphoid organs of patients with spontaneous prostate cancer regression. Front Immunol. 8:563.
Garrison E. 2012. Vcflib: a C library for parsing and manipulation VCF files. https://github.com/ekg/vcflib, last accessed September 2017.
Goessling W, McKee P, Mayer R. 2002. Merkel cell carcinoma. J Clin Oncol. 20(2):588–598.
Hamede R, McCallum H, Jones M. 2013. Biting injuries and transmission of Tasmanian devil facial tumour disease. J Anim Ecol. 82(1):182–190.
Hamede R, et al. 2015. Transmissible cancer in Tasmanian devils: localized lineage replacement and host population response. Proc R Soc B 282(1814):20151468.
Hansen TF. 2006. The evolution of genetic architecture. Annu Rev Ecol Evol Syst. 37(1):123–157.
Harold D, et al. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 41(10):1088–1093.
Hartwell L. 1992. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. Cell 71(4):543–546.
Hawkins C, et al. 2006. Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. Biol Conserv. 131(2):307–324.
Hendricks S, et al. 2017. Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). Conserv Genet. 18(4):977–982.
Hsieh P, et al. 2016. Whole-genome sequence analyses of Western Central African Pygmy hunter-gatherers reveal a complex demographic history and identify candidate genes under positive selection. Genome Res. 26(3):279–290.
Jones M, et al. 2008. Life-history change in disease-ravaged Tasmanian devil populations. Proc Natl Acad Sci U S A. 105(29):10023–10027.
Joshi N, Fass J. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle, last accessed January 2017.
Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1):27–30.
Kardos M, Husby A, McFarlane SE, Qvarnström A, Ellegren H. 2016. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. Mol Ecol Resour. 16(3):727–741.

Katsios C, Roukos D. 2010. Individual genomes and personalized medicine: life diversity and complexity. Per Med. 7(4):347–350.
Khanna C, et al. 2006. The dog as a cancer model. Nat Biotechnol. 24(9):1065–1066.
Konczal M, Babik W, Radwan J, Sadowska E, Koteja P. 2015. Initial molecular-level response to artificial selection for increased aerobic metabolism occurs primarily through changes in gene expression. Mol Biol Evol. 32(6):1461–1473.
Kumar-Sinha C, Chinnaiyan A. 2018. Precision oncology in the age of integrative genomics. Nat Biotechnol. 36(1):46–60.
Lascorz J, et al. 2010. Genome-wide association study for colorectal cancer identifies risk polymorphisms in German familial cases and implicates MAPK signalling pathways in disease susceptibility. Carcinogenesis 31(9):1612–1619.
Lazenby B, et al. 2018. Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. J Appl Ecol. 55(3):1368–1379.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.
Lotterhos K, Whitlock M. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 24(5):1031–1046.
Luo Z, Yuan C, Meng Q, Ji Q. 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. Nature 476(7361):442–445.
Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21):2957–2963.
Majewski J, Ott J. 2002. Distribution and characterization of regulatory elements in the human genome. Genome Res. 12(12):1827–1836.
Makalowska I, Lin C, Makalowska W. 2005. Overlapping genes in vertebrate genomes. Comput Biol Chem. 29(1):1–12.
Manolio T, et al. 2009. Finding the missing heritability of complex diseases. Nature 461(7265):747–753.
Margres M, et al. 2018. Large-effect loci affect survival in Tasmanian devils infected with a transmissible cancer. Mol Ecol. 00:1–11; doi:10.1111/mec.14853
Margres M, et al. 2017. Quantity, not quality: rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. Mol Biol Evol. 34(12):3099–3110.
Martin J, et al. 2004. The sequence and analysis of duplication-rich human chromosome 16. Nature 432(7020):988–994.
Mathelier A, Shi W, Wasserman W. 2015. Identification of altered *cis*-regulatory elements in human disease. Trends Genet. 31(2):67–76.
Mayor C, et al. 2000. Vista: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16(11):1046–1047.
McCallum H. 2008. Tasmanian devil facial tumour disease: lessons for conservation biology. TREE 23(11):631–637.
McCallum H, et al. 2009. Transmission dynamics of Tasmanian devils facial tumor disease may lead to disease-induced extinction. Ecology 90(12):3379–3392.
McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.
McLaren W, et al. 2016. The Ensembl variant effect predictor. Genome Biol. 17(1):122.
Metzger M, et al. 2016. Widespread transmission of independent cancer lineages within multiple bivalve species. Nature 534(7609):705–709.
Millanta F, Asproni P, Canale A, Citi S, Poli A. 2016. COX-2, mPGES-1 and EP2 receptor immunohistochemical expression in canine and feline malignant mammary tumours. Vet Comp Oncol. 14(3):270–280.
Miller W, et al. 2011. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). Proc Natl Acad Sci U S A. 108(30):12348–12353.
Missotten G, de Wolff-Rouendaal D, de Keizer R. 2008. Merkel cell carcinoma of the eyelid. review of the literature and report of patients with Merkel cell carcinoma showing spontaneous regression. Ophthalmology 115(1):195–201.
Mooney M, Nigg J, McWeeney S, Wilmot B. 2014. Functional and genomic context in pathway analysis of GWAS data. Trends Genet. 30(9):390–400.
Morandin C, et al. 2016. Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. Genome Biol. 17(1):43.
Murchison E, et al. 2012. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell 148(4):780–791.
Murchison E, et al. 2010. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science 327(5961):84–87.
Murgia C, Pritchard J, Kim S, Fassati A, Weiss R. 2006. Clonal origin and evolution of a transmissible cancer. Cell 126(3):477–487.
O'Rourke MG, Bell JR. 1986. Merkel cell tumor with spontaneous regression. J Dermatol Surg Oncol. 12(9):994–996.
Ostrander E, Davis B, Ostrander G. 2016. Transmissible tumors: breaking the cancer paradigm. Trends Genet. 32(1):1–15.
Pang C, Sharma D, Sankar T. 2015. Spontaneous regression of Merkel cell carcinoma: a case report and review of the literature. Int J Surg Case Rep. 7:104–108.
Pearse A, Swift K. 2006. Allograft theory: transmission of devil facial-tumour disease. Nature 439(7076):549.
Phan G, et al. 2003. Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. Proc Natl Acad Sci U S A. 100(14):8372–8377.
Power R, Parkhill J, de Oliveira T. 2017. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet. 18(1):41–50.
Pye R, et al. 2016. A second transmissible cancer in Tasmanian devils. Proc Natl Acad Sci U S A. 113(2):374–379.
Rokyta D, Margres M, Ward M, Sanchez E. 2017. The genetics of venom ontogeny in the eastern diamondback rattlesnake (*Crotalus adamanteus*). PeerJ 5:e3249.
Rokyta DR, Joyce P, Caudle SB, Wichman HA. 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat Genet. 37(4):441–444.
Rosenberg S, Restifo N, Yang J, Morgan R, Dudley M. 2008. Adoptive cell transfer: a clinical path to effective cancer immunotherapy. Nat Rev Cancer. 8(4):299–308.
Saadatpour A, Lai S, Guo G, Yuan G. 2015. Single-cell analysis in cancer genomics. Trends Genet. 31(10):576–586.
Schiffman J, Breen M. 2015. Comparative oncology: what dogs and other species can teach us about humans with cancer. Philos Trans R Soc Lond B Biol Sci. 370(1673): 20140231.
Sengupta N, MacFie T, MacDonald T, Pennington D, Silver A. 2010. Cancer immunoediting and "spontaneous" tumor regression. Pathol Res Pract. 206(1):1–8.
Siddle H, et al. 2013. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. Proc Natl Acad Sci U S A. 110(13):5103–5108.
Stadler BMR, Stadler PF, Wagner GP, Fontana W. 2001. The topology of the possible: formal spaces underlying patterns of evolutionary change. J Theor Biol. 213(2):241–274.
Stammnitz M, et al. 2018. The origins and vulnerabilities of two transmissible cancers in Tasmanian Devils. Cancer Cell 33(4):607–619.
Stelzer G, et al. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 1–30.
Storfer A, et al. 2018. The devil is in the details: genomics of transmissible cancers in Tasmanian devils. PLoS Pathogens. 14(8):e1007098.
Tran E, Robbins P, Rosenberg S. 2017. 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. Nat Immunol. 18(3):255–262.

Tuzun E, et al. 2005. Fine-scale structural variation of the human genome. Nat Genet. 37(7):727–732.
Verity R, et al. 2017. minotaur: a platform for the analysis and visualization of mulitvariate results from genome scans with R Shiny. Mol Ecol Resour. 17(1):33–43.
Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nat Rev Genet. 12(3):204–213.
Wei C, et al. 2006. A global map of p53 transcription-factor binding sites in the human genome. Cell 124(1):207–219.
Weir BS, Cockerham CC. 1984. Estimating *f*-statistics for the analysis of population structure. Evolution 38(6):1358–1370.
Weller M, et al. 2010. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? Nat Rev Neurol. 6(1):39–51.
Wright B, et al. 2017. Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. Sci Rep. 7(1):423.
Yang J, et al. 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 47(10):1114–1120.
Yard B, et al. 2016. A genetic basis for the variation in the vulnerability of cancer to DNA damage. Nat Commun. 7:11428.

Associate editor: Mary O'Connell